A BAYESIAN APPROACH TO UNCOVERED BACKGROUND AND MOVING PIXEL DETECTION

Kristine Matthews and Nader Namazi
Department of Electrical Engineering
The Catholic University of America
Washington, DC 20064
USA

ABSTRACT

We have formulated and evaluated a binary hypothesis test for the detection of uncovered background pixels between image frames in a noisy image sequence where we assume additive, white, gaussian noise. We have extended the binary hypothesis test to a 3-ary hypothesis test to allow for the segmentation of the image into three regions: uncovered background, stationary and moving pixels. We have evaluated both the binary and 3-ary hypothesis tests using a single measurement and multiple measurements for classifying each pixel on synthetic images, and we have evaluated the 3-ary hypothesis test on the Trevor image sequence.

PROBLEM FORMULATION

The detection of uncovered background pixels and moving pixels in image sequences is an essential part of uncovered background prediction and motion compensation for sequence coding. Many such schemes use change detection to determine uncovered background and moving portions of the image [1]. Change detection, however, does not adequately detect such regions in noisy images. In this work, we formulate and evaluate a binary hypothesis test for the detection of uncovered background pixels in noisy image sequences. We then extend the binary hypothesis test to a 3-ary hypothesis test to distinguish between stationary, moving and uncovered background pixels.

We begin by writing the intensity of the previous frame as

$$z_1(\boldsymbol{k}) = s(\boldsymbol{k}) + w_1(\boldsymbol{k}) \,, \tag{1}$$

where $\boldsymbol{k}$ denotes the spatial location of the pixel in the image frame, $z_1(\boldsymbol{k})$ and $s(\boldsymbol{k})$ are the noisy and noise-free intensities, respectively, and $w_1(\boldsymbol{k})$ is zero-mean, additive, white, gaussian noise. We can then model the noise-free intensity at each pixel in the current frame as a displaced value from the previous frame or as an uncovered background value, which leads to

$$z_2(\boldsymbol{k}) = \begin{cases} s(\boldsymbol{k}-\mathbf{d}(\boldsymbol{k})) + w_2(\boldsymbol{k}), & \boldsymbol{k} \notin \Gamma_b \\ b(\boldsymbol{k}) + w_2(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_b \,, \end{cases} \tag{2}$$

for the noisy intensity of the current frame, where $\mathbf{d}(\boldsymbol{k})$ is a nonuniform displacement vector, $b(\boldsymbol{k})$ is the intensity of the scene background, $w_2(\boldsymbol{k})$ is zero-mean, additive, white, gaussian noise and Γ_b is the region of uncovered background. Assuming that $\mathbf{d}(\boldsymbol{k})$ is small enough such that the first-order approximation of $s(\boldsymbol{k}-\mathbf{d}(\boldsymbol{k}))$ is valid and defining

$$\xi(\boldsymbol{k}) = z_2(\boldsymbol{k}) - z_1(\boldsymbol{k}) \,, \tag{3}$$

and

$$w(\boldsymbol{k}) = w_2(\boldsymbol{k}) - w_1(\boldsymbol{k}) \,, \tag{4}$$

we obtain

$$\xi(\boldsymbol{k}) = \begin{cases} -\boldsymbol{g}^T(\boldsymbol{k})\boldsymbol{d}(\boldsymbol{k}) + w(\boldsymbol{k}), & \boldsymbol{k} \notin \Gamma_b \; :H_1 \\ b(\boldsymbol{k}) - s(\boldsymbol{k}) + w(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_b \; :H_0 \,, \end{cases} \tag{5}$$

where $\mathbf{g}(\boldsymbol{k})$ is the intensity gradient vector at $\boldsymbol{k}$, and ^{T} denotes matrix transposition. H_1 and H_0 indicate the binary hypotheses, where H_1 indicates a correspondence to a stationary ($\mathbf{d}(\boldsymbol{k})=\mathbf{0}$) or moving ($\mathbf{d}(\boldsymbol{k})\neq\mathbf{0}$) pixel from the previous frame, and H_0 corresponds to a pixel belonging to an uncovered background region. Using (5) we form the likelihood ratio for the binary hypothesis test [2]

$$\Lambda[\xi(\boldsymbol{k})] \triangleq \frac{p(\xi(\boldsymbol{k})|H_1)}{p(\xi(\boldsymbol{k})|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta = \frac{P_0}{P_1} \,, \tag{6}$$

where $p(\xi(\boldsymbol{k})|H_1)$ and $p(\xi(\boldsymbol{k})|H_0)$ are the conditional probability density functions (pdf's) of $\xi(\boldsymbol{k})$ given H_1 and H_0, respectively, and η is a threshold value which depends on the *a priori* probabilities, P_1 and P_0, of H_1 and H_0 occurring, respectively. When we use a single measurement, $\xi(\boldsymbol{k})$ is the measurement we use to classify a pixel at location $\boldsymbol{k}$. When we use multiple measurements, we use the measurements at the eight nearest neighbors of $\boldsymbol{k}$ in addition to $\xi(\boldsymbol{k})$ (i.e., a 3x3 neighborhood around $\boldsymbol{k}$) to make the decision.

Following [3] and assuming that $b(\boldsymbol{k})$ and $s(\boldsymbol{k})$ are statistically independent. we write (6) as

0-7803-2431-5/95 $4.00 © 1995 IEEE

$$\Lambda[\xi(\boldsymbol{k})] = \frac{\exp\left\{\frac{1}{2}\ \mathbf{m}^T\ \mathrm{C}\ \mathbf{m}\right\} I(\xi(\boldsymbol{k}))}{2\pi |K_d|^{1/2} \int_{\Omega_q} \exp\left\{-\frac{-2\xi(\boldsymbol{k})q(\boldsymbol{k})+q^2(\boldsymbol{k})}{2\sigma_\omega^2}\right\} p(q(\boldsymbol{k}))\ dq}, \tag{7}$$

where

$$q(\boldsymbol{k}) = b(\boldsymbol{k}) - s(\boldsymbol{k}) , \tag{8}$$

$p(q(\boldsymbol{k}))$ is the pdf of $q(\boldsymbol{k})$, Ω_q is the support of $q(\boldsymbol{k})$ and σ_ω^2, K_d, **m**, C and $I(\xi(\boldsymbol{k}))$ are as in [3].

We extend (6) to a 3-ary hypothesis test to separate the non-background pixels into moving and stationary pixels by writing (2) as

$$z_2(\boldsymbol{k}) = \begin{cases} s(\boldsymbol{k}-\boldsymbol{d}(\boldsymbol{k})) + w_2(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_m \\ s(\boldsymbol{k}) + w_2(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_s \\ b(\boldsymbol{k}) + w_2(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_b , \end{cases} \tag{9}$$

where Γ_s, Γ_m and Γ_b are the regions of stationary, moving and uncovered background pixels, respectively. Defining $\xi(\boldsymbol{k})$ and $w(\boldsymbol{k})$ as in the binary case, the three hypotheses become

$$\xi(\boldsymbol{k}) = \begin{cases} -\boldsymbol{g}^T(\boldsymbol{k})\boldsymbol{d}(\boldsymbol{k}) + w(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_m \ :H_0 \\ w(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_s \ :H_1 \\ b(\boldsymbol{k}) - s(\boldsymbol{k}) + w(\boldsymbol{k}), & \boldsymbol{k} \in \Gamma_b \ :H_2 , \end{cases} \tag{10}$$

leading to the following two likelihood ratios

$$\Lambda_1[\xi(\boldsymbol{k})] \triangleq \frac{p(\xi(\boldsymbol{k})|H_1)}{p(\xi(\boldsymbol{k})|H_0)} = \frac{2\pi |K_d|^{1/2}}{\exp\left\{\frac{1}{2}\mathbf{m}^T\ \mathrm{C}\ \mathbf{m}\right\} I(\xi(\boldsymbol{k}))}$$

$$\Lambda_2[\xi(\boldsymbol{k})] \triangleq \frac{p(\xi(\boldsymbol{k})|H_2)}{p(\xi(\boldsymbol{k})|H_0)} = \frac{1}{\Lambda[\xi(\boldsymbol{k})]} . \tag{11}$$

We use the 3-ary hypothesis test as given in [2]. The test depends on P_i, the *a priori* probability of a pixel belonging to class i, and C_{mn}, the cost of deciding that a pixel belongs to class m when the pixel actually belongs to class n. Note that $\Lambda_1[\xi(\boldsymbol{k})]$ is the inverse of the likelihood ratio used in [3], and $\Lambda_2[\xi(\boldsymbol{k})]$ is the inverse of the likelihood ratio developed for the binary hypothesis case above.

EVALUATION

To evaluate the binary and 3-ary hypothesis tests, we generated several test image sequences consisting of several moving objects and different background scenes. We used both hypothesis tests on the images at several signal-to-noise ratios (SNR's) using a single measurement for classifying each pixel and multiple measurements for classifying each pixel. We define SNR as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\textit{average image variance}}{\textit{noise variance}}\right)\ dB \ . \tag{12}$$

Since the noise samples are statistically independent, for multiple measurements, $\xi_i(\boldsymbol{k})$ for i=1,...,9, the joint pdf is the product of the individual pdf's. Thus, the likelihood ratios for the multiple-measurement experiments are the products of the likelihood ratios from the single-measurement experiments.

For the binary hypothesis test, we generated receiver operating characteristics (ROC's) at several SNR's. We compared the ROC's for single measurements to those for multiple measurements. For the 3-ary hypothesis test, we formed the confusion matrix for the 20 dB SNR test case. This matrix shows the percentage of pixels of a given class that were detected in each of the three classes. We formed confusion matrices for several different cost matrices, where we define a cost matrix as a matrix, **C**, with entry m, n equal to C_{mn}, where m=0,1,2 and n=0,1,2.

In addition to evaluating the 3-ary hypothesis test on synthetic images, we evaluated it on the Trevor image sequence. We visually inspected the resulting image segmentation as our performance criterion.

All of the above tests require knowledge of the pdf of the difference between background and object intensity, $p(q(\boldsymbol{k}))$, in the region of uncovered background. We determine this pdf by convolving the histogram of the background with the *flipped* histogram of the object and normalizing to make the result a valid pdf. Given that a likely application of this technique for image sequence coding is video teleconferencing, we made the following assumptions. If the application is video teleconferencing, knowledge of the background is readily available. The background is the scene before the respective conference attendee enters the staging area. We therefore assume that we have knowledge of the intensity distribution of the background. We also assume that the moving object is likely to be a person, and again we have statistical knowledge of the object.

For the tests involving the Trevor sequence, we interactively extracted the background region from a noisy frame of the sequence different from the present or previous frame and used that region to obtain the necessary statistics. In some experiments, we used a frame of the sequence that was not our present or previous frame to extract the object for the object statistics. In other experiments, we used a frame from the Miss America sequence, and interactively extracted the person to use for the object statistics. There were no visual distinctions in the segmentations resulting from using Miss America or Trevor for the intensity histogram of the object.

Another approach, given the same video teleconferencing application, is to take the first image frame and consider the middle one third as the object and the two outer thirds as the background in order to obtain the required histograms. This assumption is reasonable since the person is most likely to be centered in the image frame.

RESULTS

In Figures 1-5 and Table 1, we summarize the results of the experiments done on the synthetic test images. Figure 1 shows the previous and present frames of one of the test image sequences at an SNR of 20 dB. Figure 2 shows a segmentation of the present frame (Γ_s - black, Γ_m - white, Γ_b - gray) resulting from the 3-ary hypothesis test using single measurements for pixel classification and two different cost matrices, C_1 and C_2, where

$$C_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad C_2 = \begin{bmatrix} 0 & 1 & 1 \\ 9 & 0 & 9 \\ 1 & 1 & 0 \end{bmatrix} . \qquad (13)$$

From this figure, we see that if we make it more costly to misclassify a moving or uncovered background pixel as a stationary pixel, we increase the probability of detection of uncovered background and moving pixels, but we also misclassify many of the stationary pixels.

Figure 3 illustrates the effectiveness of multiple measurements in the low SNR situation. The figure shows a segmentation of a 10 dB image using a single measurement for pixel classification and multiple measurements for pixel classification. Figure 4 is the true segmentation for this test sequence. In Figure 5a, we show the ROC's for several SNR's and single measurements, and in Figure 5b, we compare the ROC's at 10 dB for single and multiple measurements. The ROC's in Figure 5 indicate the increased detection performance gained by using multiple measurements in low SNR images. Table 1 is the confusion matrix resulting from the 3-ary hypothesis test on the 20 dB test sequence. From this table, we see the effect of the cost matrix on the performance of the hypothesis test.

Figures 6-8 show the results of the 3-ary hypothesis test using the Trevor image sequence. Figure 6 shows the noisy (20 dB) present and previous frames. In Figures 7 and 8, we show segmentations of the 20 dB sequence and 5 dB sequence resulting from single and multiple measurements for pixel classification, respectively using $C=C_1$. These figures also demonstrate the advantage of using multiple measurements in low SNR sequences. The segmentations of the Trevor sequence are good with the exception that the tie is consistently misclassified. This occurs because the histogram of the tie is similar to the histogram of the background.

References:

[1] R. Thoma and M. Bierling, "Motion Compensating Interpolation Considering Covered and Uncovered Background," *Signal Processing: Image Communication 1*, 1989, pp. 191-212.

[2] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley, 1968.

[3] N. M. Namazi, P. Penafiel and C. M. Fan, "Nonuniform Image Motion Estimation Using Kalman Filtering," *IEEE Transactions on Image Processing*, Sept. 1994, Vol. 3, No. 5, pp 678-683.

[4] N. M. Namazi, *New Algorithms for Variable Time Delay and Nonuniform Image Motion Estimation*, Ablex Publishing Corporation, 1994.

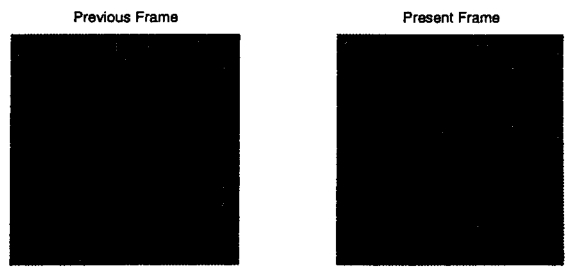

Figure 1: Test image sequence at SNR=20 dB

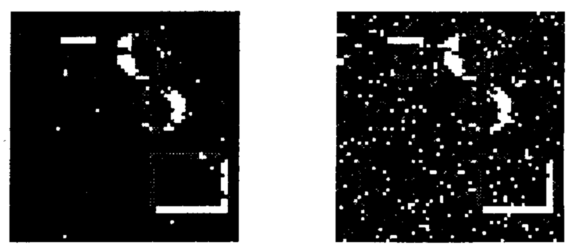

Figure 2: Segmentation of 20 dB image for different C (Γ_s - black, Γ_m - white, Γ_b - gray) left: $C=C_1$ right: $C=C_2$

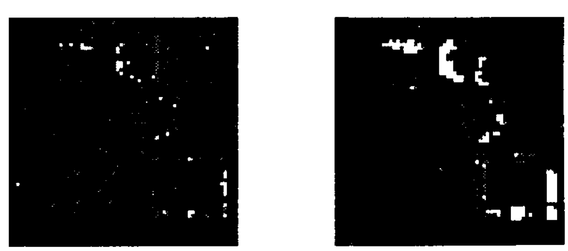

Figure 3: Segmentation of 10 dB image using C_1 (Γ_s - black, Γ_m - white, Γ_b - gray)
left: single measurements right: multiple measurements

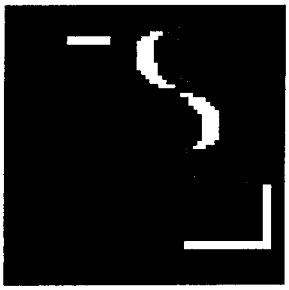

Figure 4: True segmentation
(Γ_s - black, Γ_m - white, Γ_b - gray)

DETECTED REGION	TRUE REGION		
	moving	stationary	uncovered background
moving	84 / 89	0.3 / 6	6 / 7
stationary	8 / 3	99 / 89	13 / 4
uncovered background	8 / 8	0.3 / 5	81 / 89

Table 1. Confusion matrix, in percentage, for 3-ary hypothesis test (single measurements, SNR=20 dB, $\mathbf{C}_1$ / $\mathbf{C}_2$)

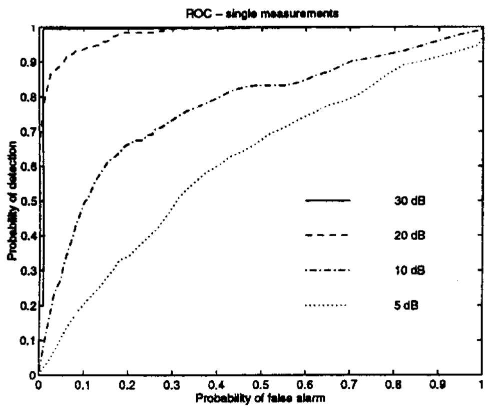

Figure 5a: ROC single measurements

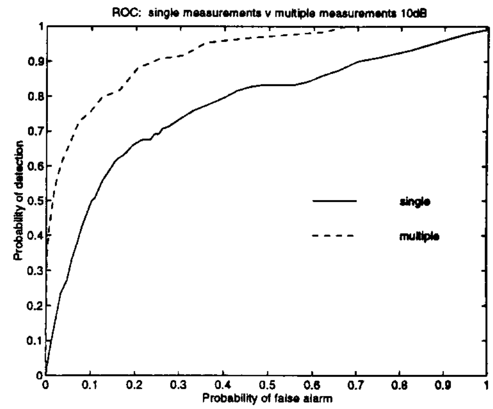

Figure 5b: ROC, single measurements and multiple measurements for SNR=10 dB

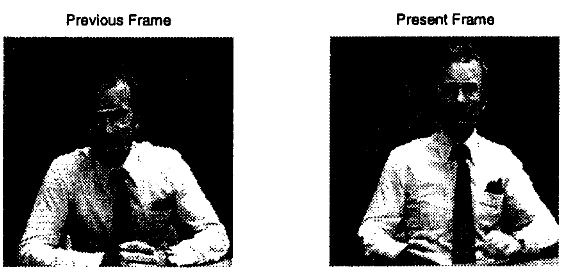

Figure 6: Trevor frames SNR=20 dB

Figure 7: Segmentation from single measurements using $\mathbf{C}_1$ (Γ_s - black, Γ_m - white, Γ_b - gray) left: 20 dB right: 5 dB

Figure 8: Segmentation from multiple measurements using $\mathbf{C}_1$ (Γ_s - black, Γ_m - white, Γ_b - gray) left: 20 dB right: 5 dB